AI for Research

MCAST  
Malta College of Arts, Science & Technology

Email: ramon.falzon.a100947@mcast.edu.mt

Table of Contents

[Introduction 3](#_Toc42112838)

[What is Artificial Intelligence? 3](#_Toc42112839)

[Applications of AI 3](#_Toc42112840)

[Data Collection 4](#_Toc42112841)

[Transformation of text into numbers 5](#_Toc42112842)

[Methodology 6](#_Toc42112843)

[Results 7](#_Toc42112844)

[Discussion 10](#_Toc42112845)

[Bibliography 11](#_Toc42112846)

# Introduction

## What is Artificial Intelligence?

AI (Artificial Intelligence) is generally considered to be a vague term by many, but this holds no weight. This is because AI is confused with ML (Machine Learning); whereas AI is a script that handles inputs, processes these inputs, and then react depending on these inputs, ML is what makes the script learn. AI can be as simple as a simple card-game playing script or even as complex as a recommendation system. However, it should be noted that AI is hard-coded, so it is a pre-defined script with pre-defined answers. If the AI is wrong, it will require a developer to fix the script. What ML adds to an AI requires an AI to be created with ML in mind, as the code required for an AI to support ML is far different then an AI without ML. ML operates by adjusting the inputs by their respective weights (A number that represents the importance of that weight); the weights alter themselves depending on how close they were to the expected output and then re-try. However, the altering of the weights only comes in when the Model is in the process of being developed, not after it is developed. A model is what an AI with ML is called after it has been trained enough until its weights are accurate enough. After a model is trained, an accuracy rating will be available, this accuracy rating judges either how far off the AI can be from the correct answer or the probability of being wrong when a model is given data to give predictions off of. It should be noted that the “probability” of being wrong is not a probability of chance, but how reliable the model is when it is given data that is not as easily recognizable as regular data. (Jarek & Mazurek, 2019)

## Applications of AI

AI has a very wide plateau of applications in almost every field; one of the more explored fields of applications for AI is the field of marketing. In a study conducted in 2019 by Krystyna Jarek and Grzegorz Mazurek, applications of AI in marketing are: Voice processing technologies *(Using e-shopping to buy products via voice-activated software or executing tasks such as setting timers or adjusting integrated home-appliances via voice software or dedicated hardware such as Siri)*, text processing technologies *(Augmented reality which provides a virtual assistant as you navigate around shopping centers, high lighting what you are after or giving an explanation of what you are looking at and another type of augmented reality which not only acts as a GPS but doubles as a touring guide, recommending point of interests that the user might be interested in),* Image recognition and processing technology *(Facial recognition replaces passwords when making payments with your e-banking card or application or image recognition which tries to analyze the object in the photo and then search for information about it or related products),* decision making (Product recommendation such as Netflix’s recommendation system or Amazon) and in automated robotics and vehicles (Inventory stock taking robots & service free shops). (Jarek & Mazurek, 2019)

Furthermore, in a study conducted by H.Isah, P.Trundle & D.Negau, it was proven that a model could be trained to observe social media, understand what the people are commenting about products and then label them accordingly. (Isah, et al., 2014)

## Data Collection

It is argued that during the data collection process, none of it should be discarded. Removal of data from the dataset should only be considered when the dataset is far too large, and even then, removing data should be done at random to avoid any prejudice or bias.

It is mentioned that during data collection, data can be grouped up into their own categories to be tested on its own or for categorical reasons. Categories can also come in the forms of hierarchies. It is suggested that the more specific the category, the more accurate the predictions will be for that category. However, this is not necessary; only when a specific sub-category is desired or for data presentation purposes.

(J. Srnka & T. Koeszegi, 2007)

## Transformation of text into numbers

Transforming Qualitative data, such as user-written reviews composed of text, into Quantitative data, a numerical value that represents a rating in the given example, is not a very simple process. There are multiple methodologies that can be used to transform qualitative data into quantitative data. One method by Xueying (2019), was to first pre-process the data, removing any words that the computer can not recognize and then divide the number of positive words by the total number of words in a sentence. One error with this method occurred due to the pre-processing of data; by having deleted some words, some sentences ended up with no words thus creating invalid numbers that are above 1 (Which should be impossible). The pre-processing methodology was used to tag the data for the model in Xueying’s study. (Xueying, 2019)

# Methodology

This paper will closely follow a tutorial provided by visualstudiomagazine (James, 2018). The tutorial makes use of the Python 3 programming languages, Keras, Tensorflow and a set of movie reviews from IMDB. The format of the reviews are as follows: Limited to a number of words, have already been tagged and if the number of words in the review are less than the limit, empty padding will be added.

Three models will be created; the first model will be an exact re-creation of the model in the tutorial, the second model will have slightly altered variables as to produce more accurate results to see if the model can do better. The variables will be set as follows: 25,000 unique words (Changed from the suggested 20,000), the model will limit reviews to have a maximum of 300 words (Changed from the suggested 80) and the embedded vector length will be 300 (, while the third model will be a very serious attempt at creating a very accurate model that can match expectations. In the third model, the variables will be set as follows: the amount of unique words will be set to 40,000 words (Changed from the suggested 20,000), the model will limit reviews to have a maximum of 1,000 words in it (Changed from the suggested 80) and the embedded vector length will be 500 (The higher bound of suggested common use).

Each model will be locally saved, and a Python 3 script that can make use of the models will be created. The Python 3 script will ask the user which version of the model the user wants to use & then prompt the user to input a review. Afterwards, the model will try to predict what the review-score would have been with the given review.

The models will be tested with various types of data, such as: Invalid use, non-existent words, professional reviews, regular reviews, reviews that use slang words, and reviews with mis-spelled words.

# Results

The results given were very desirable, as most of the predictions given by the model were what close to or even more accurate than would be expected. There were some cases where the model failed to predict correctly, in one specific case where the review was “This film was mediocre at best”, the model predicted that the would-be scored would have been even worse than a review stating “This film is terrible”; this most likely happened due to the word mediocre not being known by the model.

Text A: "I did not like this at all, do not watch this." (regular reviews)

Text B: "I really like this film! It is a must watch for everybody. I really loved the script of the film and all the emotions it brings, I cried when the dog died, and I laughed when I realized it was just playing dead! Chris J really did a good job on this film, his acting was a masterpiece, second to none!" (Professional reviews)

Text C: "This film is mediocre at best." (regular reviews)

Text D: "This film is not good or bad." (regular reviews)

Text E: "great great great great great great great great great great great great great great great great" (Invalid use)

Text F: "I’d rather have my eyes removed than be forced to watch this again" (regular reviews)

Text G: "bad bad bad bad bad bad bad bad bad bad bad bad bad bad bad bad bad bad bad bad bad bad bad bad bad bad bad" (Invalid use)

Text H: “Fo sure the best film of 2019!! None better than this!” (Slang words)

Text I: “Th best film of 2020, I can not belive how good this is. Watch it!” (Spelling mistakes)

Text J: “This film is so bad, it makes the sound of ‘badom’ every time you watch it…” (Non-existent words)

Text K: “This film on the whole is great, but there are some flaws that one can does not ignore. Often, in the background, the cameras recording the film can be seen. It is not easy to spot them, but once you notice one, you start noticing more. The acting was not the best, but it was not that bad either. Overall, I still recommend you watch it.” (Professional reviews)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Expected | Model 1 | Deviancy | Model 2 | Deviancy | Model 3 | Deviancy |
| Text A | 0.30 | 0.15 | 0.15 | 0.30 | 0 | 0.01 | 0.29 |
| Text B | 0.90 | 0.66 | 0.24 | 0.83 | 0.07 | 0.83 | 0.07 |
| Text C | 0.50 | 0.14 | 0.36 | 0.74 | 0.24 | 0.02 | 0.48 |
| Text D | 0.50 | 0.50 | 0 | 0.46 | 0.04 | 0.09 | 0.41 |
| Text E | 1.00 | 1.00 | 0 | 1.00 | 0 | 1.00 | 0.00 |
| Text F | 0.10 | 0.23 | 0.13 | 0.13 | 0.03 | 0.30 | 0.20 |
| Text G | 0.00 | 0.00 | 0 | 0.00 | 0 | 0.00 | 0.00 |
| Text H | 0.75 | 0.62 | 0.13 | 0.53 | 0.22 | 0.63 | 0.12 |
| Text I | 0.75 | Error |  | 0.53 | 0.22 | 0.26 | 0.49 |
| Text J | 0.15 | 0.67 | 0.52 | 0.78 | 0.63 | 0.65 | 0.50 |
| Text K | 0.75 | 0.11 | 0.64 | 0.47 | 0.28 | 0.05 | 0.70 |

Average Deviancy for Model 1: (0.15 + 0.24 + 0.36 + 0 + 0 + 0.13 + 0 + 0.13 + 0.52 + 0.64) / 10 = 0.217

Average Deviancy for Model 2: (0 + 0.07 + 0.24 + 0.04 + 0 + 0.03 +0 + 0.22 + 0.22 + 0.63 + 0.28) / 11 = 0.157273

Average Deviancy for Model 3: (0.29 + 0.07 + 0.48 + 0.41 + 0 + 0.2 + 0 +0.12 + 0.49 + 0.50 + 0.70) / 11 = 0.296364

In Text I, the script throws an error for Model 1 because of the word “believe” being written as “belive”. In Model 2 & 3, this error did not occur because during their creation, their unique word count was higher than Model 1.

# Discussion

Before the beginning of the discussion, it should be noted that the test sample was not very large, nor did it judge the models depending on each types of review, only one collective group that has a variety of possibilities. However, given the level of generic the reviews were, a conclusion can still be drawn.

The first model, the one suggested by the tutorial, was not too far from expectations. On the simpler reviews, it was generally not too far from expectations. In review C, the review that makes use of the word ‘mediocre’, the model was not close to the expected outcome, far enough to be considered as wrong.

As for the second model, the model with slight improvements over the first model, it performed generally well all around, with many predictions being near spot on or spot on. It also had issues with the word ‘mediocre’, it was not as far off but the prediction is still far enough to be considered as wrong. As for Text K, one of the reviews which would need a more accurate prediction over the others, it still failed to produce a relevant prediction. It should be noted however, that Model 2 produced over all desirable results.

Lastly, the third model, the model with larger training data sets, more epochs, and a larger vocabulary. Model Three required a large amount of time to train, due to its larger processing load. Despite these facts, the produced model, as bulky as it is, failed to produce good predictions. Almost all the predictions created by the model were inferior to other less trained models. On the simpler of the reviews, the model failed to produce accurate results. The reason why the more trained model has produced worse result than a model with a smaller and more limited training set might be due to a learning curve of accuracy vs data set, where having a small data set allows a model to create more accurate results on simpler testing data, while opening up the model to more words and larger reviews would require a number of data that is substantially larger than its predecessors.

# Bibliography

Isah, H., Trundle, P. & Neagu, D., 2014. *Social media analysis for product safety using text mining and sentiment analysis.* [Online]   
Available at: https://ieeexplore.ieee.org/abstract/document/6930158  
[Accessed 03 06 2020].

J. Srnka, K. & T. Koeszegi, S., 2007. *From Words to Numbers: How to Transform.* [Online]   
Available at: https://www.dphu.org/uploads/attachements/books/books\_3754\_0.pdf  
[Accessed 3 6 2020].

James, M., 2018. *visualstudiomagazine.* [Online]   
Available at: https://visualstudiomagazine.com/articles/2018/11/01/sentiment-analysis-using-keras.aspx  
[Accessed 5 28 2020].

Jarek, K. & Mazurek, G., 2019. *Marketing and Artificial Intelligence.* [Online]   
Available at: https://www.researchgate.net/publication/333671063\_Marketing\_and\_Artificial\_Intelligence  
[Accessed 5 26 2020].

Xueying, W., 2019. *Using Sentiment Analysis for Comparing Attitudes between Computer Professionals and Laypersons on the Topic of Artificial Intelligence.* [Online]   
Available at: https://dl.acm.org/doi/abs/10.1145/3342827.3342829  
[Accessed 03 06 2020].